Random proportional Weibull hazard model for large-scale information systems

Kobayashi, Kiyoshi; Kaito, Kiyoyuki

RANDOM PROPORTIONAL WEIBULL HAZARD MODEL FOR LARGE-SCALE INFORMATION SYSTEMS

Structured Abstract

In this study, the purpose of which is to aid in the asset management of large-scale information systems supporting infrastructure, a Weibull hazard model is used to formulate the failure generation process of component wear-out failure, the rate of which changes with time. Information systems are composed of many devices. In order to consider the heterogeneity of the hazard rate of each device, the random proportional Weibull hazard model, which expresses the heterogeneity of the hazard rate as random variables, is proposed. Furthermore, the authors develop a methodology that expresses the heterogeneity of hazard rates in gamma distribution as well as estimates unknown parameters and the heterogeneity of hazard rates contained in the hazard model. Finally, using historical data regarding actual failures in the traffic control system of expressways, the authors estimate the wear-out failure rate of components. The validity of the methodology is investigated through a case study. Concretely, as for HDD which mainly composes information systems, the service life at which the survival probability is 50% is estimated as 158 months. However, even for the same HDD, use environment differs according to usage. Actually, among the 3 different usages (PC, server, others), failures happen earliest in the case of PCs, which have the highest heterogeneity parameter and a survival probability of 50% after 135 months of usage. On the other hand, as for others, its survival probability is 50% at 303 months.

RANDOM PROPORTIONAL WEIBULL HAZARD MODEL FOR LARGE-SCALE INFORMATION SYSTEMS

ABSTRACT. In this study, the purpose of which is to aid in the asset management of large-scale information systems supporting infrastructure, a Weibull hazard model is used to formulate the failure generation process of component wear-out failure, the rate of which changes with time. Information systems are composed of many devices. In order to consider the heterogeneity of the hazard rate of each device, the random proportional Weibull hazard model, which expresses the heterogeneity of the hazard rate as random variables, is proposed. Furthermore, the authors develop a methodology that expresses the heterogeneity of hazard rates in gamma distribution as well as estimates unknown parameters and the heterogeneity of hazard rates contained in the hazard model. Finally, using historical data regarding actual failures in the traffic control system of expressways, the authors estimate the wear-out failure rate of components. The validity of the methodology is investigated through a case study.

Keywords; asset management, large-scale information system, deterioration prediction, Random proportional Weibull hazard model, historical data, heterogeneity

INTRODUCTION

In order to achieve efficiency of operation and speedy provision of information to users of large-scale infrastructure, information systems have been developed which are composed of various monitoring sensors and processing/output devices. Asset management of infrastructure is an important issue as well. Points to be noted for asset management of information systems are as follows: 1) that the information system is a large-scale system formed from an enormous number of components, 2) that the system has a hierarchical structure whereby a failure in an individual device can possibly develop into a functional failure of the whole system, and 3) that the deterioration of the information system's function, such as through obsolescence or physical deterioration, is an important management point.

When performing asset management of information systems, it is necessary to consider the differences in management at the 1) component level, 2) system level, and 3) function level, with the latter points requiring a more integrated consideration. Among these, the authors develop a random proportional Weibull hazard model to carry out a fault analysis at the component level, of the components in large-scaled information systems. Of course, in order to perform asset management of information systems, a fault analysis at the level of each component is not enough, and it is necessary to develop a methodology that simultaneously achieves the above three points. The fault analysis model for the component level proposed in this study can be a basic analytical tool for constructing integrated asset management systems of large-scale information systems.

Information systems are composed of various types of components. These components can be divided into two groups: the accidental failure rate of equipment, which is unrelated to time, and the wear-out failure rate, which increases over time. Generally, regarding the fault process of accidental equipment failure, the hazard rate expressed by the moment's failure rate density can be formulated by an exponential hazard model that does not depend on time (or a Poisson based model of failure events). On the other hand, with wear-out equipment failure, a non-homogeneous hazard model that takes into consideration the hazard rate's dependency on time is necessary. With wear-out equipment failure, there is the characteristic that the failure rate increases as time progresses after the system is installed. Therefore, in order for decisions to be made about renewals and replacements, information regarding the deteriorating process of wear-out equipment failure is needed. For this reason, the authors focus on wear-out equipment failure, using a Weibull hazard model, known as a representative non-homogeneous hazard model, to analyze the mechanism of the failure rate increasing over time. However, large-scale information systems comprise many components, and different types of components might have different hazard rates. Therefore, when analyzing faults of information systems that comprise various types of devices and components, it is important to consider the heterogeneity of the hazard rates that exist between the different types of components. In this study, with this in consideration, the random proportional Weibull hazard model, whose heterogeneity of hazard rates is subject to a gamma distribution, is formulated and a methodology is proposed which estimates the failure rate of various components comprising an information system.

BASIC IDEAS OF THIS STUDY

Overview of Existing Studies

In traditional hazard analysis (Cox and Oakes, 1984, Lancaster, 1990), the targeted system is assumed to be constructed of elements of the same quality, and the aim is to express a model of failure generation randomly attained according to a certain hazard function. With hazard analysis, the random failure generation process is modeled, and a deterministic model called a hazard function is employed. In reality, large-scale information systems, such as the one targeted for actual analysis in this study, have complex structures composed of an enormous number of many types of components. Managing and operating a large-scale information system requires important decisions regarding policy issues such as when to change individual components and how components should be stored. However, the failure rate of all the components might not be expressed by the same hazard rate. Rather, it is natural to assume that the hazard rate differs for each type of component. As for methods expressing the heterogeneity of the hazard rates of different types of components, it is possible to 1) express the difference of the component characteristics in dummy variables or to 2) consider the probability distribution of the hazard rate. The former method is simple and easy to understand. However, as the number of types of components increases, the number of dummy variables needed to express the components' characteristics also increases, causing the efficiency of the model's estimation results to decline considerably. In fact, in large-scale information systems, even components of the same type, categorized by type and device, have different deterioration characteristics depending on how they are used and where they are located, and these systems are structures with extremely divided deterioration characteristics. In hazard models, it is not practical to use dummy variables for deterioration characteristics sub-divided in this way. Therefore, in this study the authors choose the latter method, using a random proportional Weibull hazard model that expresses the heterogeneity of hazard rates of different components in probability distribution, and formulating a model for the failure process of component groups that compose information systems.

There have been many studies on the hazard analysis considering the heterogeneity of hazard rate. Especially, there have been an enormous amount of researches into the mixed hazard model in which the heterogeneity of hazard rate exists for each sample (Cameron and Trivedi, 1990, 1998, Cruz, 2002, Gourieroux and Visser, 1986, McNeil, et al., 2005, Mikosch, 2000). In the mixed hazard model, it is assumed that the heterogeneity parameters that determine the hazard function is subject to a certain probability density function. The mixed hazard model is defined with the probabilistic convolution of the hazard function and the probability distribution of heterogeneity parameters. The ordinary Poisson process has the condition that the average of the probabilities of rare events is equal to the variance of them. In this situation, the mixed Poisson process model has been researched, in order to increase the degree of freedom in describing the variance of probabilities in the Poisson process. Kaito and Kobayashi (2008) applied it to asset management, by producing a model of the arrival process of road obstacles. For analyzing the failure of a device that breaks down accidentally, it is effective to use the mixed Poisson process model that reflects the heterogeneity of failure rate. However, monitors, input and output devices, and CPUs, etc., which constitute information systems, are the devices that break down due to deterioration, and so it is necessary to take into account the heterogeneity of Weibull hazard rate. In this study, the author proposes a random proportional Weibull deterioration hazard model, which describes the heterogeneity of Weibull hazard rate with the gamma distribution. The gamma distribution is a general exponential probability distribution, including an exponential distribution, and has the characteristics that can describe a broad range of probability distributions. In addition, the probabilistic convolution of the Weibull distribution and the gamma distribution is simple, and it is possible to analytically derive the random proportional Weibull deterioration hazard model. Accordingly, it can be considered that this model has practically excellent features.

Take in Figure (1)

Basic Frame of Modeling

This is a model for the occurrence and process of failure events in information systems. As Figure 1 shows, the components in information systems are in three layers: 1) Type, 2) Device, and 3) Components. The type-layer contains hard disk drive (HDD), power supply, and processing and monitoring equipment. An information system is composed of M-unit Type components and each component is represented by the suffix i ($i=1, ..., M$). Type i components are used for N_i-unit Device and each Device is represented by the suffix j ($j=1, ..., N_i$). In a traffic control system, for example, each Type of component is used in different Devices such as personal computers (PC), servers, and so on. Since each Device has its own application of components, its failure probability is different from that of others. For Device j ($j=1, ..., Ni$), L_{ij}-unit Type i components are used and each component is represented by the suffix k ($k=1, ..., L_{ij}$). Components in each Type and each Device are considered to have different hazard rates. But the failure process in components in each Device is considered to be described by using the same hazard rates. Here, an infinitely continuous time axis starting from time point t=0 is used. If the existing information systems as a whole are renewed at t=0, the deterioration of each component starts from t=0. When a component has a failure, it is immediately replaced. The new one is supposed to have the same performance that the old one had. Now, take a look at $t=T$ where a certain period of time has passed. Then, a failure history is obtained as shown in Figure 1, which cites an example of a failure history of Device

2 (server). Device 2 is composed of L_2-unit components. Among them, component A has no failure from t=0. The time of use of component A is T and the lifetime of component A is considered to be longer than T. On the other hand, component B had a failure twice at T_1 and T_2. The first lifetime $\zeta = T_1$ and the second $\zeta = T_2 - T_1$.

Here, it is supposed that each type of component has the failure characteristics of the wear-out failure component. With the wear-out failure component, as shown in Figure 2, the generation rates of failures (hazard rates) grow as the elapsed time moves farther from the nearest time of renewal. This type of lifespan distribution for the wear-out failure component is often used to express the components' time-dependent deterioration. It is assumed that this is subject to the Weibull distribution. Furthermore, the hazard rates of different types of components can be expressed in time functions, as are shown in Figure 2. Functions like this that express the hazard rate's temporal change are called hazard functions. The hazard functions of each component are in an expansion or contraction by a factor of a baseline hazard function. A model that expresses proportionally expanded or contracted functions is called a proportional hazard model. If the failure process of each type of component that composes each device can be mutually expressed with a proportional hazard model, the heterogeneity in hazard rates can be expressed by the probability distribution of proportionality constants of hazard functions. Information systems comprise many devices, but in most cases, the number of the same type of components in each device is not that large. By estimating the parameter of the standard proportional Weibull hazard function and the parameter of the probability distribution that expresses the heterogeneity of the proportionality constant between the types, the random proportional Weibull hazard model can easily express the heterogeneity of the hazard rates between types and components. On the other hand, if the heterogeneity of the Weibull hazard rate cannot be expressed by the proportional hazard model, it becomes necessary to estimate the Weibull hazard model for the different types and devices. However, if the number of components of the same type that compose each device is small, it becomes difficult to estimate the Weibull hazard model. If the above points are considered, the random proportional Weibull hazard model proposed in this study is very effective for expressing the failure process of information systems that have sub-divided component structures.

Take in Figure (2)

RANDOM PROPORTIONAL WEIBULL HAZARD MODEL

Formulation of Random Proportional Weibull Hazard Model

Large-scale information systems are, as shown in Figure 1, composed of M types of components, and the component at number i (i = 1, …, M) is used in a total of N_i devices. Furthermore, the total number of type i components used in device j is L_{ij}. Among the type i components, the component at number k (k = 1, …, L_{ij}) out of the components that compose device j (j = 1, …, N_i) is picked up. The elapsed time since this component has been renewed will be expressed as t_{ij}^k. The arrival rate of each component's failure occurrence is based on the random proportional Weibull hazard model.

$$\lambda_{ij}(t_{ij}^k) = \varepsilon_{ij}\gamma_i m(t_{ij}^k)^{m-1} \qquad (1)$$

However, γ_i is the parameter expressing the arrival density of type *i*, and *m* is the acceleration parameter. Equation (1) is the general Weibull hazard function (Aoki *et al*, 2007; Tsuda *et al*, 2006) with the parameter ε_{ij} (hereinafter, heterogeneity parameter) expressing the heterogeneity (Maher, 1996) of the hazard rates of type *i* and device *j*. The heterogeneity parameter expresses the heterogeneity of the hazard rates between the components of different devices and types. Especially, in the case that $\varepsilon_{ij} = 1$, the random proportional Weibull hazard function (1) matches the general Weibull hazard function. This kind of hazard function is called a baseline hazard function (See Figure 2.). However, for the same component used in the same device, the heterogeneity parameter has a common value. The heterogeneity parameter takes a deterministic value in reality, but is a parameter that is impossible to observe. Also, the probability density function $f_{ij}(t_{ij}^k)$ and survival probability $\overline{F}_{ij}(t_{ij}^k)$ of the lifespan of type *i* component *k* in device *j* are as follows.

$$f_{ij}(t_{ij}^k) = \varepsilon_{ij}\gamma_i m (t_{ij}^k)^{m-1} \exp\left\{-\gamma_i \varepsilon_{ij}(t_{ij}^k)^m\right\} \tag{2a}$$

$$\overline{F}_{ij}(t_{ij}^k) = \exp\left\{-\gamma_i \varepsilon_{ij}(t_{ij}^k)^m\right\} \tag{2b}$$

The value of the heterogeneity parameter is a probabilistic variable which is subject to a certain probability distribution. The random proportional Weibull hazard model (1) has the same deterioration acceleration parameter *m* for all components, but the hazard arrival density $\varepsilon_{ij}\gamma_i m$ expresses a proportionately different deterioration characteristic for each type and device. In this study, a Weibull hazard model in which the hazard arrival density has a proportional distribution for each type and device, is defined as a random proportional Weibull hazard model.

Here, the heterogeneity parameter ε_{ij} is subject to the gamma distribution. Furthermore, use the case in which there are different averages of the heterogeneity parameter for each type. Gamma distribution as a special form contains exponential distribution, and can express the exponential probability density function family defined by $[0,\infty)$. Also, it has the merit that it is easy to handle analytically. Here, the parameter γ_i expresses the average hazard arrival density of the type *i* component, and that the heterogeneity parameter ε_{ij} is a probability error term that is subject to the gamma distribution of average 1, variance ϕ^{-1}. The gamma distribution is defined by $[0,\infty)$, and regarding the arbitrary explanatory variable and probability error term, the right side of equation (1) is guaranteed to have a positive value. Generally, the probability density function $g(\varepsilon_{ij} : \alpha, \beta)$ of the gamma distribution $G(\alpha, \beta)$ can be defined as follows.

$$g(\varepsilon_{ij} : \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \varepsilon_{ij}^{\alpha-1} \exp\left(-\frac{\varepsilon_{ij}}{\beta}\right) \tag{3}$$

The average of gamma distribution $G(\alpha, \beta)$ is $\mu = \alpha\beta$, and the variance is $\sigma^2 = \alpha\beta^2$. Also, $\Gamma(\ )$ is the gamma function. Furthermore, the probability density function $\overline{g}(\varepsilon_{ij} : \phi)$ of the gamma distribution of average 1, variance ϕ^{-1} can be expressed as follows.

$$\overline{g}(\varepsilon_{ij} : \phi) = \frac{\phi^\phi}{\Gamma(\phi)} \varepsilon_{ij}^{\phi-1} \exp(-\phi\varepsilon_{ij}) \tag{4}$$

Estimating Method of the Model

The random proportional Weibull hazard model has unknown parameters such as arrival density parameters γ_i (i=1,…, M) for each type, an acceleration parameter *m*, different heterogeneity parameters ε_{ij} (i = 1, …, M ; j = 1, …, N_i) for each type and device and the variance parameter ϕ of the heterogeneity parameter. Normally with Weibull hazard models, the parameters γ_i and m can be estimated from failure history data. However, with the random proportional Weibull hazard model, it is necessary to estimate the variance parameter ϕ of the heterogeneity parameter, and the heterogeneity parameters ε_{ij} (i = 1, …, M ; j = 1, …, N_i) for each type and device, as well. For reader's convenience, the proposed estimation flow is shown in Fugre 3 with the following detail dscription.

Take in Figure (3)

Now the database of the failure history of the targeted system is available. All the information regarding the time that each component failed (was exchanged) since the time of installation, of the targeted system, is stored in the database. Then, the failure history of the component is expressed as $\Xi = (\xi_1,\ldots,\xi_M)$, and $\xi_i = (\xi_{i1},\ldots,\xi_{iNi})$ is the failure history of type i component. Also, ξ_{ij} is the failure history for the type i component of device j, and is $\xi_{ij} = \{(\delta_{ij}^1, t_{ij}^1),\ldots,(\delta_{ij}^{Lij}, t_{ij}^{Lij})\}(i = 1,\ldots,M,\ j = 1,\ldots,N_i)$. Also, δ_{ij}^k is the dummy variable which takes the value 1 if the type i component k (k =1, …, L_{ij}) of device j fails, and the value 0 if it does not. t_{ij}^k is the in-service time (or lifespan) of the type i component k of device j. Therefore, when $\delta_{ij}^k = 0$, t_{ij}^k is the length of time since installation to the present. On the other hand, when $\delta_{ij}^k = 1$, t_{ij}^k is the lifespan. Let us discuss it in detail with the servers A and B (device 2) of the power source (type 2) shown in Figure 1. Server A has been used continuously from the installation of the system to the present T without failure or replacement. Accordingly, $(\delta_{22}^A = 0, t_{22}^A = T)$ is recorded in the database as the information on the history of failure regarding Server A. In this study, the information on whether there is any failure and usage time is considered as a basic information unit. On the other hand, Sever B broke down twice at T_1 and T_2 and was replaced since the installation of the system. Accordingly, as the information on the history of failure regarding server B, $(\delta_{22}^{B1} = 1, t_{22}^{B1} = T_1)$, $(\delta_{22}^{B2} = 1, t_{22}^{B2} = T_2 - T_1)$, and $(\delta_{22}^{B3} = 0, t_{22}^{B3} = T - T_2)$ are recorded. Obviously, even if several devices are observed for the same period of time, the basic information unit varies according to the history of failure. When it is possible to gain several basic information units from a single device like server B, we should discriminate each basic information unit by adding a superscript to B like δ_{22}^{B1}, but let us omit this superscript notation.

Here it is assumed that the heterogeneity parameter $\bar{\varepsilon}_{ij}$ is given. The conditional likelihood $\ell_{ij}(\xi_{ij} : \gamma_i, m, \bar{\varepsilon}_{ij})$ with observed data ξ_{ij} concerning faults of type i components of device j can be expressed as follows:

$$\ell_{ij}(\xi_{ij} : \gamma_i, m_i, \bar{\varepsilon}_{ij}) = \prod_{k=1}^{Lij} \left\{ \bar{F}_{ij}(t_{ij}^k : \gamma_i, m_i, \bar{\varepsilon}_{ij}) \right\}^{(1-\delta_{ij}^k)} \left\{ f_{ij}(t_{ij}^k : \gamma_i, m_i, \bar{\varepsilon}_{ij}) \right\}^{\delta_{ij}^k}. \tag{5}$$

However, with the above equation, the probability density function of the lifespan distribution $\bar{f}_{ij}(t_{ij}^k : \gamma_i, m_i, \bar{\varepsilon}_{ij})$ and the survival function $\bar{F}_{ij}(t_{ij}^k : \gamma_i, m_i, \bar{\varepsilon}_{ij})$ are explicitly expressed as the parameter $\gamma_i, m_i, \bar{\varepsilon}_{ij}$ function. Here, when the heterogeneity parameter ε_{ij} is subject to the standard gamma distribution $\bar{g}(\varepsilon_{ij} : \phi)$, the likelihood function of the observed data ξ_{ij} can be expressed as follows:

$$\begin{aligned} L_{ij}(\xi_{ij} : \theta_i) &= \int_0^\infty \prod_{k=1}^{Lij} \left\{ \bar{F}_{ij}(t_{ij}^k : \gamma_i, m_i, \bar{\varepsilon}_{ij}) \right\}^{(1-\delta_{ij}^k)} \left\{ f_{ij}(t_{ij}^k : \gamma_i, m_i, \bar{\varepsilon}_{ij}) \right\}^{\delta_{ij}^k} \bar{g}(\varepsilon_{ij} : \phi) d\varepsilon_{ij} \\ &= \frac{\phi^\phi}{\Gamma(\phi)} \prod_{k=1}^{Lij} \left\{ \gamma_i m (t_{ij}^k)^{m-1} \right\}^{\delta_{ij}^k} \int_0^\infty \varepsilon_{ij}^{s_{ij}+\phi-1} \exp\left\{ -(\phi + \gamma_i \tau_{ij}) \varepsilon_{ij} \right\} d\varepsilon_{ij}. \end{aligned} \tag{6}$$

However, $\theta = (\gamma_i, m_i, \phi)$. Also, $s_{ij} = \sum_{k=1}^{Lij} \delta_{ij}^k, \tau_{ij} = \sum_{k=1}^{Lij} (t_{ij}^k)^{m-1}$. With the above equation, the heterogeneity parameter ε_{ij} of all type i components of device j takes the same value. Notice that to express this, the authors define the likelihood function $L_{ij}(\xi_{ij} : \theta_i)$ as the expected value regarding the probability variable ε_{ij} of the conditional likelihood $\ell_{ij}(\xi_{ij}:\gamma_i, m_i, \bar{\varepsilon}_{ij})$. Hence, by the transformation of the variable $x_{ij} = \varepsilon_{ij}(\phi + \gamma_i \tau_{ij})$, the following is obtained:

$$L_{ij}(\xi_{ij}:\theta_i) = \frac{\phi^\phi}{\Gamma(\phi)} \prod_{k=1}^{Lij} \left\{ \gamma_i m (t_{ij}^k)^{m-1} \right\}^{\delta_{ij}^k} \int_0^\infty \left(\frac{x_{ij}}{\phi + \gamma_i \tau_{ij}} \right)^{s_{ij}+\phi-1} \exp(-x_{ij}) \frac{dx_{ij}}{\phi + \gamma_i \tau_{ij}}$$

$$= \frac{\phi^\phi}{\Gamma(\phi)} \frac{\Gamma(s_{ij}+\phi)}{(\phi + \gamma_i \tau_{ij})^{s_{ij}+\phi-1}} \prod_{k=1}^{Lij} \left\{ \gamma_i m (t_{ij}^k)^{m-1} \right\}^{\delta_{ij}^k} . \qquad (7)$$

Therefore, the logarithmic likelihood function with the observed data $\Xi = (\xi_1, ..., \xi_M)$ can be expressed as

$$\ln L(\Xi, \theta) = \sum_{i=1}^{M} \sum_{j=1}^{Ni} \ln L_{ij}(\xi_{ij}:\theta_i) = N\phi \ln \phi - \sum_{i=1}^{M} \sum_{j=1}^{Ni} (s_{ij}+\phi) \ln(\phi + \gamma_i \tau_{ij})$$

$$+ \sum_{i=1}^{M} \sum_{j=1}^{Ni} \sum_{k=0}^{s_{ij}-1} \ln(\phi + k) + \sum_{i=1}^{M} \sum_{j=1}^{Ni} \sum_{k=1}^{Lij} \delta_{ij}^k \left\{ \ln \gamma_i + \ln m + (m-1) \ln t_{ij}^k \right\} . \qquad (8)$$

However, each element of $\theta = (\theta_1, \ldots, \theta_{M+2})$ is expressed as $(\theta_1, \ldots, \theta_M) = (\gamma_1, \ldots, \gamma_M)$, $\theta_{M+1} = m$, $\theta_{M+2} = \phi$. Also, $N = \sum_{i=1}^{M} N_i$, and with the third item on the right side of equation (8), when $s_{ij} = 0$, $\sum_{k=0}^{s_{ij}-1} \ln(\phi + k) = 0$ is defined. Also, when s_{ij} =1, $\sum_{k=0}^{s_{ij}-1} \ln(\phi + k) = \ln \phi$.

The maximum likelihood estimator of the parameter value $\boldsymbol{\theta}$, which maximizes the logarithmic likelihood function (8), is given as $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_{M+2})$ which simultaneously satisfies

$$\frac{\partial \ln L(\Xi, \theta)}{\partial \theta_i} = 0, \quad (i = 1, \ldots, M+2) \cdot \qquad (9)$$

Furthermore, the estimated value $\hat{\Sigma}(\hat{\theta})$ of the asymptotic covariance matrix of the parameter can be expressed as follows:

$$\hat{\Sigma}(\hat{\theta}) = \left[\frac{\partial^2 \ln L(\Xi, \hat{\theta})}{\partial \theta \partial \theta'} \right]^{-1} . \qquad (10)$$

However, the inverse matrix of the right side of the above equation is the inverse matrix of the Fisher information matrix after $(M+2) \times (M+2)$ with elements $\partial^2 \ln L(\Xi, \hat{\theta}) / \partial \theta_i \partial \theta_j$. The parameter's maximum likelihood estimator can be obtained by solving the non-linear simultaneous equation (9) of M+2 dimension. In this study, the maximum likelihood estimator is estimated by the Newton-Raphson method. If the maximum likelihood estimator $\hat{\theta}$ is obtained, the covariance matrix estimated value $\hat{\Sigma}(\hat{\theta})$ is employed to estimate the *t*-test statistics.

Next, with the maximum likelihood estimator $\hat{\theta}$ of the parameter vector as postulate, the maximum likelihood estimator of the heterogeneity parameter ε_{ij} (i=1, …, M; j =1, …, N_i) is obtained. Here, the partial likelihood function is defined as follows:

$$L_{ij}^\circ(\xi_{ij}, \varepsilon_{ij}:\hat{\theta}_i) = \frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \prod_{k=1}^{Lij} \left\{ \hat{\gamma}_i \hat{m} (t_{ij}^k)^{\hat{m}-1} \right\}^{\delta_{ij}^k} \varepsilon_{ij}^{s_{ij}+\hat{\phi}-1} \exp\left\{ -(\hat{\phi} + \hat{\gamma}_i \hat{\tau}_{ij}) \varepsilon_{ij} \right\} . \qquad (11)$$

However, $\hat{\tau}_{ij} = \sum_{k=1}^{Lij} (t_{ij}^k)^m$. At this time, the maximum likelihood estimator of the heterogeneity parameter ε_{ij} can be obtained as $\hat{\varepsilon}_{ij}$ that satisfies

$$\frac{\partial L_{ij}^\circ(\xi_{ij}, \varepsilon_{ij}:\hat{\theta}_i)}{\partial \varepsilon_{ij}} = 0 \cdot \qquad (12)$$

The maximum likelihood estimator of the heterogeneity parameter obtained in this way is the estimated value obtained with the parameter $\hat{\theta} = (\hat{\gamma}_i, \hat{m}_i, \hat{\phi})$ as postulate. To express this explicitly, the solution of equation (12 is expressed as $\hat{\varepsilon}_{ij}(\hat{\theta}_i)$. Finally, from equation (11), (12), the following equation can be obtained:

$$\hat{\varepsilon}_{ij}(\hat{\theta}_i) = \frac{s_{ij} + \hat{\phi} - 1}{\hat{\phi} + \hat{\gamma}\hat{\tau}_{ij}}. \tag{13}$$

EMPIRICAL STUDY

Overview of Empirical Study

Targeting the traffic control system managed by a highway company, the random proportional Weibull hazard model is estimated. The traffic control system is a system that has been sequentially renewed from the old system since 1990, and has been in operation continuously. It comprises 9 central station systems (hereinafter, stations), manages the conditions of expressways, and provides appropriate real-time information to users. The operational conditions of the traffic control system are also under real-time surveillance, and in the case of faults, the failure-generated component will be specified and the time and content of the failure will be recorded.

In this study, out of all the components that compose the traffic control system, the authors targeted the component group that, in case of a fault, has the possibility of developing into a serious functional failure of the whole system, and estimated the random proportional Weibull hazard model. After investigating the failure history database of the traffic control system and interviewing the system manager, the authors selected a component group to target for model estimation, ultimately deciding on three types, HDDs (Hard disk drive), power supply, and processing devices. In the current traffic control system, there are 177 HDDs, 306 power supply, and 180 processing devices. From the same research and interview, HDDs and processing devices are used as three different devices: PCs (used as monitors and terminals), servers (used as servers and processing equipment), and other devices (other uses not included in the former two). Furthermore, these components are used in 9 different stations, each type being categorized into 27 categories.

Estimation Results

In this study, 3 types of components: HDD, power supply and processing device are used. Therefore, 3 arrival density parameters γ_i (i=1,2,3) are introduced. Also, assuming the hazard rates of the PCs, servers and other devices in the 9 stations are heterogeneous, heterogeneity parameters ε_{ij} (i=1,2,3; j=1, …, N_i) are defined for each device. Therefore, the unknown parameters to be estimated are arrival density parameter γ_i (i=1,2,3), acceleration parameter m, heterogeneity dispersion parameter ϕ, ε_{ij}. The random proportional Weibull hazard model estimated by the method proposed is shown in Table 1. The value in parentheses shows the *t*-value, and the *t*- value of either parameter; as a result, the null hypothesis that they have no explanatory power for each explanatory variable model, is rejected at significance level 0.95. It is possible to show more clearly the difference in hazard rates for each type of component. As shown in Table 1, the maximum likelihood estimator of the acceleration parameter is $\hat{m} = 2.174$. From Equation (2b), it can be seen that the survival probability of each type of component gradually decreases as the in-service time increases. Generally, if $\hat{m} = 1.00$, it would be an accidemtal failure component whose failure rate does not depend on time, but all the components used in this study have characteristics of the wear-out failure component.

Take in Table (1)

Analysis Results

The authors categorized the 3 types of components (HDD, power supply, and processing device) according to which of the 9 stations of the traffic control system they are installed in, and further subdivided them according to the 3 types of usage, 1) PCs (used as monitors, terminals), 2) Servers (used as servers, processing equipment) and 3) Other devices (Used in other ways other than described in the former two). As shown in Table 2, this creates 27 categories for each type. From equation (13), the maximum estimator of heterogeneity parameters for each category can be identically obtained. However, some categories (i,j) have no applicable components, and the number of heterogeneity parameters to be estimated are these: 21 for HDD, 9 for power supply, and 26 for processing device. Table 2 shows each of the estimated heterogeneity parameters $\hat{\varepsilon}_{ij}(\hat{\theta}_i)$ and the number of samples of each component of each type. The same table shows that the maximum estimators of heterogeneity parameters are distributed variously depending on the device, and it can be seen that in order to express the deterioration characteristic of components in an information system it is necessary to consider the heterogeneity of hazard rates. Also, because the component structure is extremely subdivided, it can be surmised that the estimation efficiency will fall if the different component characteristics with dummy variables are expressed. By using the maximum estimator of heterogeneity parameters, we can obtain a deterministic Weibull hazard model that expresses the deterioration characteristic of each device. However, as shown in Table 2, there are some device categories with a very small number of samples. Therefore, there is the possibility of a problem occurring with the reliability of the Weibull hazard models obtained for each device category. In fact, of the heterogeneity parameters shown in Table 2, there are some for which the null hypothesis that it has no explanatory power for implementing the heterogeneity parameter could not be rejected, at a significance level of 95%. Therefore, by grouping the device categories, aggregative Weibull hazard models that express the average deterioration characteristic for each group can be obtained. Let's define the heterogeneity parameters for each type and device ε_{ij}, as aggregative average heterogeneity parameters for each type $E\varepsilon_i$ (i=1, …, M). The aggregative partial likelihood functions for all devices are defined as

$$L_i^{\infty}(\xi_i, E\varepsilon_i : \hat{\theta}_i) = \frac{\hat{\phi}^{\hat{\phi}}}{\Gamma(\hat{\phi})} \prod_{j=1}^{Ni} \prod_{k=1}^{Lij} \left\{ \hat{\gamma}_i \hat{m} (t_{ij}^k)^{\hat{m}-1} \right\}^{\delta_{ij}^k} E\varepsilon_{ij}^{s_{ij}+\hat{\phi}-1} \exp\left\{ -(\hat{\phi} + \hat{\gamma}_i \hat{\tau}_{ij}) E\varepsilon_{ij} \right\}. \tag{19}$$

The maximum estimator of the heterogeneity parameter $E\varepsilon_i$ satisfies

$$\frac{\partial L_i^{\infty}(\xi_{ij}, E\varepsilon_i : \hat{\theta}_i)}{\partial E\varepsilon_i} = 0 \tag{20}$$

and can be expressed as

$$E\hat{\varepsilon}_i(\hat{\theta}_i) = \frac{\sum_{j=1}^{Ni} s_{ij} + \hat{\phi} - 1}{\sum_{j=1}^{Ni} \hat{\phi} + \hat{\gamma}\hat{\tau}_{ij}}. \tag{21}$$

Similarly, the maximum estimators $E\hat{\varepsilon}_i(\hat{\theta}_i)$ of heterogeneity parameters $E\varepsilon_{il}$ (i,l=1,2,3) aggregated for each type and usage (PC, server, others) is

$$E\hat{\varepsilon}_{il}(\hat{\theta}_i) = \frac{\sum_{j\in\omega l} s_{ij} + \hat{\phi} - 1}{\sum_{j\in\omega l} \hat{\phi} + \hat{\gamma}\hat{\tau}_{ij}}. \tag{22}$$

However, l is the usage of the components, and l=1 means PC, l=2 means servers, and l=3 means others. Also, ω_l is the set of devices of the usage l. The maximum estimators of the average heterogeneity parameters aggregated with the above idea are shown in Table 3. If the average heterogeneity parameters aggregated by type are compared, $E\hat{\varepsilon}_1(\hat{\theta}) > E\hat{\varepsilon}_3(\hat{\theta}) > E\hat{\varepsilon}_2(\hat{\theta})$ is true. Furthermore, the same relation applies to the arrival density parameter, so the hazard rate of HDD is highest, and that of power supply is smallest. By obtaining the average heterogeneity parameter $E\hat{\varepsilon}_i(\hat{\theta})$, which is the heterogeneity parameters of each type of component aggregated, the Weibull hazard model that expresses the average deterioration characteristic of each type of component can be obtained. By using the average heterogeneity parameter $E\hat{\varepsilon}_i(\hat{\theta})$ (i=1,2,3) aggregated from each type, the average survival function of each type is obtained, and these are shown in Figure 4. Of all the samples, the shown survival probabilities are the relative ratio of the samples that survived in the targeted period. The lifespan of components is generally evaluated as service life. The same figure can be illustrated to show the service life according to the survival probability. Therefore, by using the survival probability as a management indicator, the service life can be evaluated by an arbitrary management indicator. However, the management indicator should be prepared with the importance of components in mind. From Figure 4, it can be seen, for example, that the use period (service life) at which the survival probability is 50% is 158 months for HDD, 804 months for power supply, and 332 months for processing device. Furthermore, the survival probability of power supply used for 120 months is 98.9%, and it is 95% at 240 months. For processing device, it is 92.7% at 120 months and 71.1% at 240 months. It can be seen that the failure rate increases as use time lengthens for power supply and processing device as well as for HDD. However, among the 3 types, HDD has the most rapid incline, while power supply has the gentlest. Furthermore, the survival functions of HDD and processing device obtained by aggregating heterogeneity parameters by usage are shown in Figure 5 and Figure 6. Figure 5 shows the survival function of HDD. Even for the same HDD, use environment differs according to usage. As seen from the figure, among the 3 different usages (PC, server, others), failures happen earliest in the case of PCs, which have the highest heterogeneity parameter and a survival probability of 50% after 135 months of usage. On the other hand, devices of other uses have the smallest heterogeneity parameter, and the survival probability is 50% at 455 months. In Figure 6, of processing parts, the hazard rate when they are used as servers is the highest, with a survival probability of 50% at 303 months. The hazard rate is smallest for PCs, with a survival rate of 50% at 356 months.

Take in Table (2)

Take in Table (3)

Take in Figure (4)

Take in Figure (5)

Take in Figure (6)

CONCLUSIONS

In this study, a deterioration failure estimation model for components in information systems is proposed, aiming to improve asset management at the component level of large-scale information systems which support infrastructure. The authors focus on the point that information systems are composed of many types of components, and point out the necessity for a fault analysis model that can express the heterogeneity of hazard rates of different types. To operationally express the heterogeneity of failure rates, the Weibull hazard model is employed as a base, and a random proportional Weibull hazard model expressing the proportional heterogeneity of hazard rates with a standard gamma distribution is formulated. Furthermore, through a case study using a traffic control system for expressways, the validity of the proposed hazard model is empirically verified. As mentioned above, by using the random proportional hazard model considering the heterogeneity in the deterioration process, it is possible to improve the decision-making process regarding deterioration prediction for facilities and equipment, including infrastructure, and asset management based on deterioration prediction. The following are the findings of this study:

- The random proportional Weibull hazard model was formulated, in order to take into account the heterogeneity of hazard rate of each device for large-scale information system composed of a variety of component. The authors proposed a two-step estimation method using observational data and indicated that heterogeneity parameters can be calculated identically.
- The random proportional Weibull deterioration hazard model was applied to a traffic control information system for expressways, and examined the appropriateness of the proposed method. Actually, three types of devices: HDD, power supply, and processing device were classified into 27 categories (9 sections, 3 usages), and a hazard model estimation was conducted while considering each heterogeneity.
- The results of estimation focused on the difference among device types indicate that expected lifespan is 158 months for HDD, 804 months for the power source, and 332 months for the processing part. It was also found that expected lifespan varies from 135 to 455 months for HDD and from 303 to 356 months for the processing part according to the difference in purpose.

Applying the random proportional Weibull hazard model to asset management shows that there are issues requiring future study. First, it is necessary to develop an asset management methodology for components, using the proposed hazard model. Especially, there are cases in which it is difficult to acquire necessary equipment if an information component goes out of production during the operation of an information system. It is expensive to use a different component as a substitute, so that in order to avoid the problem of out of stock components, it is necessary to store replacement components. Or, to carry out preventative maintenance on information equipment, it is necessary to decide logically on the renewal time of components. The hazard model proposed in this study can be used for performing asset management at the component level. Secondly, fault analysis is necessary at the function level of information systems. With asset management at the function level, it is necessary to focus on the seriousness of the effect a failure in one component or component group can have on the functional level of the whole system, and it is necessary to consider the maintenance strategy of each component and component group. The authors are in the process of developing a methodology to analyze the system's dynamic failure characteristics, which expresses the failure process of each component using the hazard model proposed in this study, and expresses the impact of each failure on the function of the whole system using a fault-tree. Thirdly, it is necessary to work on asset management at the system level of information systems. For this, it is necessary to consider simultaneously the technical obsolescence of the information system, the delay in processing time, and the dynamic failure process, and to develop a real-time option model to determine the best timing for renewing the information system. Fourth, the proposed method is applied to asset management of other types of infrastructure, for instance pavement, since the effect of the heterogeneity of individual sections of pavement would be larger than that of information systems.

REFERENCES

Aoki, K., Yamamoto, K. and Kobayashi, K. (2007). "Optimal Inspection and Replacement Policy using Stochastic Method for Deterioration Prediction", *The 11th World Conference on Transportation Research*, Berkeley, USA

Cameron, A.C. and Trivedi, P.K. (1990). "Regression-based tests for overdispersion in the Poisson model", Journal of Econometrics, Vol.46, pp.347-34.

Cameron, A.C. and Trivedi, P.K. (1998). "Regression Analysis of Count Data", Cambridge University Press.

Cox, D. R. And Oakes, D. (1984). "Analysis of Survival Data", Chapman & Hall.

Cruz, M.G. (2002). "Modelling, Measuring and Hedging Operational Risk", Wiley.

Gourieroux, C. and Visser, M. (1986). "A count data model with unobserved heterogeneity", *Journal of Econometrics*, Vol.79, pp.247-268.

Kaito, K. and Kobayashi, K. (2008). "Obstacle Emergence Risk and Road Patrol Frequency", *Proceeding of the International Seminar on Asset Management for Developing World*, pp.42-55.

Lancaster, T. (1990). "The Econometric Analysis of Transition Data", *Cambridge University Press*.

Lawless, J.F. (1987). "Negative binomial and mixed Poisson regression", Canadian Journal of Statistics, Vol.15, pp.209-225.

Maher, M.J. and Summersgill, I. (1996). "A comparative methdology for the fitting predictive accident models", Accident Analysis & Prevention, Vol.28, pp.281-296.

McNeil, J.A., Frey, R. and Embrechts, P. (2005). "Quantitative Risk Management", Princeton University Press.

Mikosch, T. (2000). "Non-Life Insurance Mathematics", Springer.

Piegorsch, W.W. (1990). "Maximum likelihood estimation for the negative binomial dispersion parameter", *Biometrika*, Vol.46, pp.863-867.

Tsuda, Y., Kaito, K., Aoki, K. and Kobayashi, K. (2006). "Estimating Markovian Transition Probabilities for Bridge Deterioration Forecasting", *Journal of Structural Eng./Earthquake Eng., JSCE*, Vol.23, No.2, pp.241s-256s.

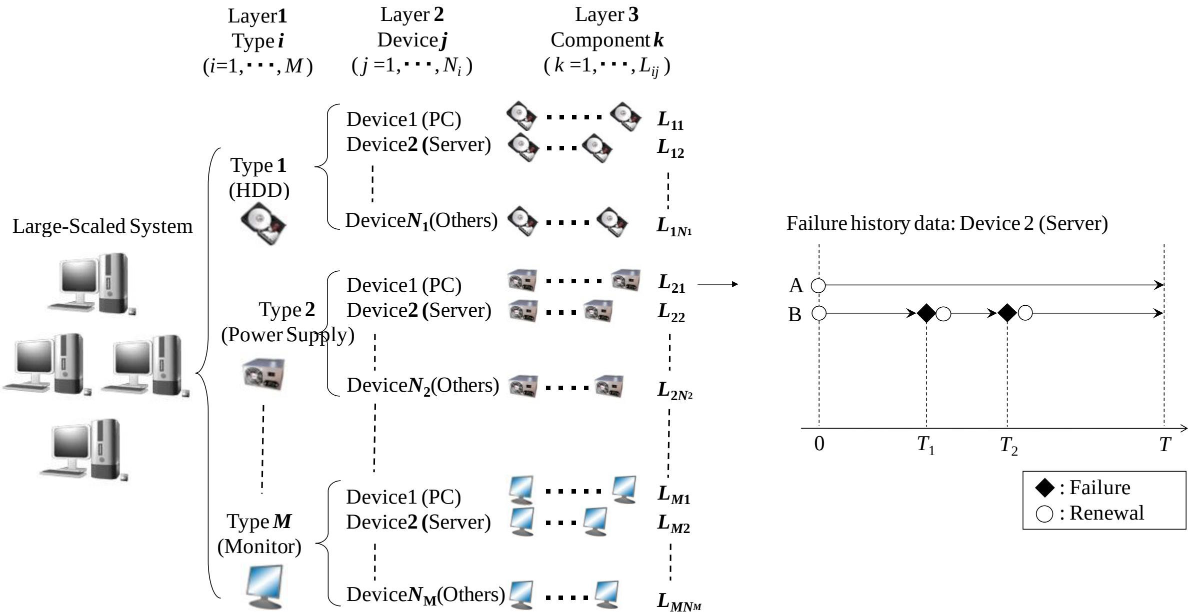
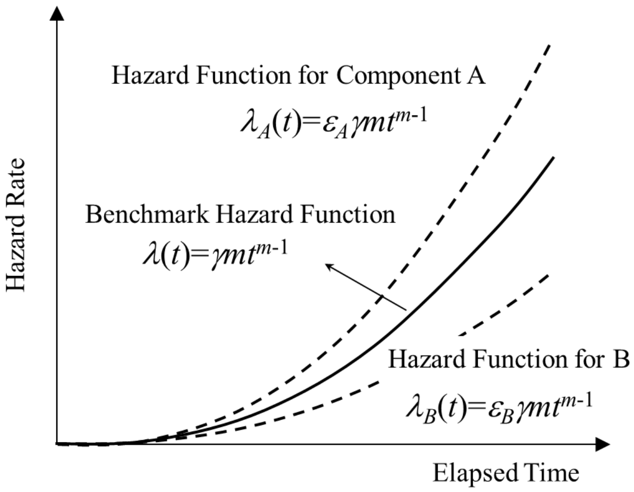


Figure 1. Large-Scaled Infromation System and History Data of Failure Occurrences



Note: The baseline model applies to ε=1. Also, the hazard function of component B is a function multiplied by $\varepsilon_B > 1$ on the baseline model, and the hazard function shifts upwards. While component is $\varepsilon_A < 1$ and the hazard rate proportionally shifts downwards.

Figure 2. Heterogeneity of Hazard functions

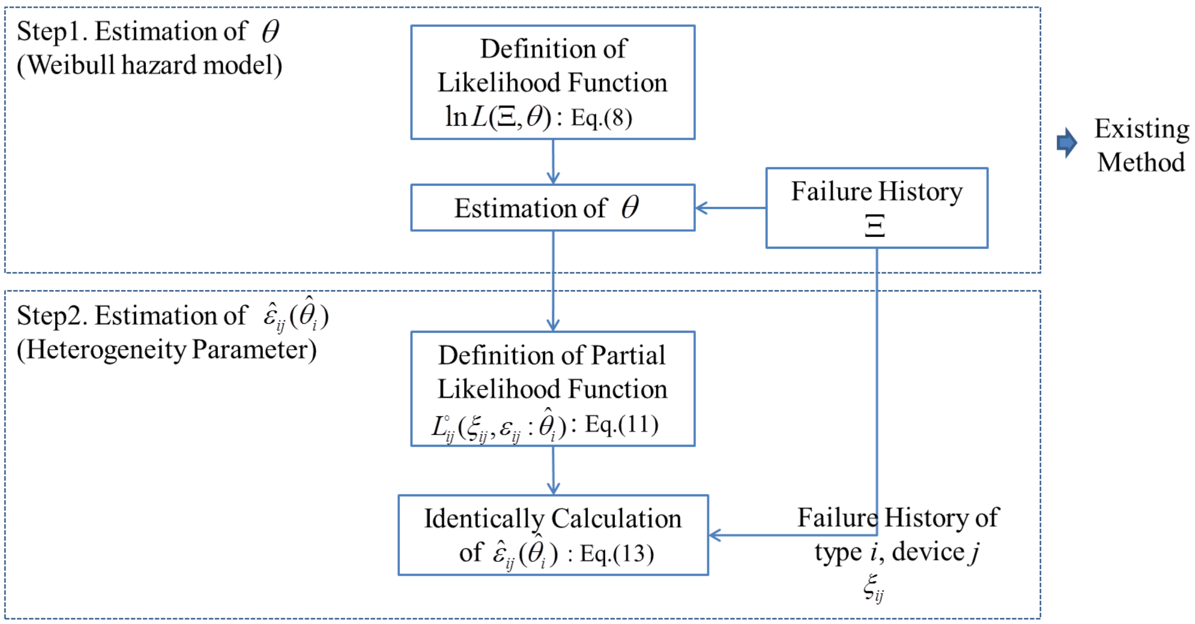


Figure 3. Estimation Fllow

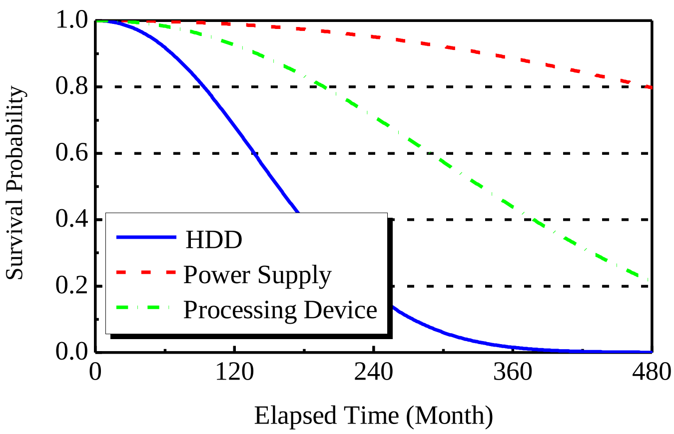


Figure 4. Survival Probabilities of Each Type

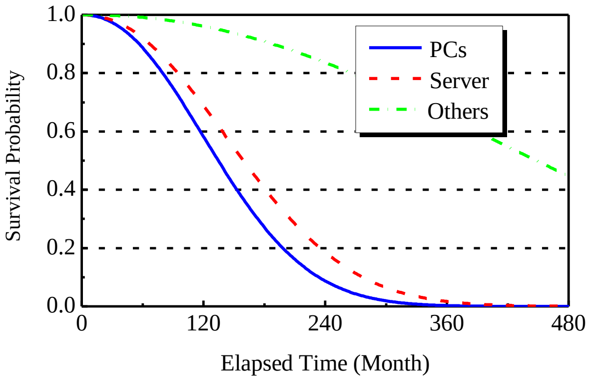


Figure 5. Survival Probabilities of HDD

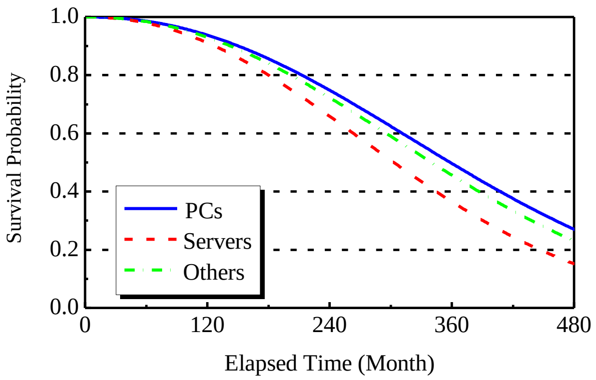


Figure 6. Survival Probabilities of Processing Device

Table 1. Estimation Results

Parameter		Model
γ	γ_1	1.251E-5 (-5.104E6)
	γ_2	1.631E-6 (-2.311E7)
	γ_3	5.293E-6 (-9.182E6)
m		2.174 (49.031)
ϕ		1.193 (2.182)
Logarithmic likelihood		-402.441

Table 2. Heterogeneity Parameters

		HDD	Power supply	Processing device
Station 1	PC	0.154 (1.536)	0.006 (2.401)	0.154 (1.398)
	Server	0.148 (1.744)	-	0.148 (1.484)
	Others	-	-	0.008 (3.694)
Station 2	PC	0.123 (2.973)	0.120 (2.586)	0.770 (8.94E-02)
	Server	2.208 (0.541)	-	0.125 (1.897)
	Others	0.161 (1.321)	-	0.008 (3.313)
Station 3	PC	0.146 (1.799)	0.007 (5.549)	0.146 (1.507)
	Server	0.669 (3.357)	-	0.008 (3.715)
	Others	0.133 (2.337)	-	0.860 (7.62E-02)
Station 4	PC	1.437 (5.38E-02)	0.004 (3.174)	0.688 (0.190)
	Server	0.768 (2.136)	-	0.674 (0.213)
	Others	0.006 (11.207)	-	0.833 (7.82E-02)
Station 5	PC	0.753 (0.479)	0.008 (1.983)	0.113 (2.170)
	Server	0.600 (1.500)	-	0.628 (0.303)
	Others	-	-	-
Station 6	PC	0.134 (2.310)	0.008 (1.954)	0.132 (1.752)
	Server	0.114 (3.365)	-	0.802 (6.24E-02)
	Others	-	-	0.142 (1.579)
Station 7	PC	0.147 (1.779)	0.007 (2.090)	0.147 (1.500)
	Server	1.304 (0.246)	-	0.136 (1.674)
	Others	-	-	0.009 (3.070)
Station 8	PC	5.400 (12.044)	1.360 (3.519)	0.481 (0.830)
	Server	1.833 (0.181)	-	1.508 (0.325)
	Others	-	-	0.581 (0.424)
Station 9	PC	0.844 (0.178)	0.632 (3.44E-02)	0.416 (1.260)
	Server	0.138 (2.140)	-	0.948 (3.43E-03)
	Others	-	-	0.123 (1.937)

Table 3. Heterogeneity parameters

	HDD	Power supply	Processing device
$E\hat{\varepsilon}_i\left(\hat{\theta}\right)$	0.923 (9.746)	0.205 (8.737)	0.431 (20.086)
$E\hat{\varepsilon}_{i1}\left(\hat{\theta}\right)$	1.302 (0.022)	-	0.366 (8.344)
$E\hat{\varepsilon}_{i2}\left(\hat{\theta}\right)$	0.900 (6.403)	-	0.527 (3.973)
$E\hat{\varepsilon}_{i3}\left(\hat{\theta}\right)$	0.095 (14.552)	-	0.410 (8.243)

RANDOM PROPORTIONAL WEIBULL HAZARD MODEL FOR LARGE-SCALE INFORMATION SYSTEMS

Dean, Professor, Dr. Kiyoshi KOBAYASHI
Graduate School of Management, Kyoto Univ.
kkoba@psa.mbox.media.kyoto-u.ac.jp
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

1978, M. A., Kyoto Univ.
1978-1986, Research Associate, Kyoto University
1987-1990, Assoc. Professor, Dept. of Social Systems Eng., Tottori Univ.
1990-1996, Professor, Dept. of Social Systems Eng., Tottori Univ.
1996-present, Professor, Graduate School of Engineering, Kyoto Univ.
2006-2009, Vice Dean, Professor, Graduate School of Management, Kyoto Univ.
2007-present, Professor, Research Institute of Management Studies, Kyoto Univ.
2010-present, Dean, Professor, Graduate School of Management, Kyoto Univ.

Associate Professor, Dr. Kiyoyuki KAITO
Frontier Research Center, Graduate School of Eng., Osaka Univ.
kaito@ga.eng.osaka-u.ac.jp
2-1 Yamada-Oka, Suita, Osaka, 565-0871, Japan

2000, Ph.D, Dept. of Civil Eng., The Univ. of Tokyo
2000-2001, Visiting Researcher, Dept. of Civil Eng., The Univ. of Tokyo
2001, Visiting Scholar, Dept. of Civil Eng. and Eng. Mechanics, Columbia Univ.
2001-2002, Postdoctral Fellow, Dept. of Civil Eng., The Univ. of Tokyo
2002-2004, R&D Manager, Bridge Maintenance Consultant Corp.
2004-2006, Senior Researcher, Bridge Maintenance Consultant Corp.
2006-2007, Visiting Researcher, Dept. of Urban Management, Kyoto Univ.
2007-present, Assoc. Professor, Frontier Research Center, Graduate School of Eng., Osaka Univ.